

DOCUMENT RESUME

ED 274 204

FL 016 088

AUTHOR Li, Wen-hui
TITLE Research on the Cloze Procedure as Measurement of Reading Comprehension and Its Use in the EFL Classroom.
PUB DATE 86
NOTE 11p.; In: WATESOL Working Papers, Number 3, Spring 1986; see FL 016 085.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cloze Procedure; *English (Second Language); *Language Tests; *Reading Comprehension; Second Language Instruction; Test Construction; Test Reliability; *Test Use; Test Validity

ABSTRACT

The paper reviews the ways in which the cloze testing procedure measures the reading comprehension skills of students in the English as a foreign language (EFL) classroom. The construction and scoring of the test and the interpretation of the scores are discussed and the reliability and validity of the cloze procedure for testing reading skills are explored. Evidence suggests that the cloze procedure is a more reliable testing tool in the EFL classroom than the traditional reading comprehension multiple choice examination. In general, the cloze test is a practical and powerful measurement. Although more research of the tests must be done, the cloze reading comprehension test, a component of testing, can play an important role and is expected to be used more extensively by EFL teachers.
(TR)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

RESEARCH ON THE CLOZE PROCEDURE AS MEASUREMENT OF READING
COMPREHENSION AND ITS USE IN THE EFL CLASSROOM

Wen-hui Li
Georgetown University

ED 274 204

The cloze procedure as a measuring technique is no longer really new. It was first introduced by W.L. Taylor in 1953, but it attracted little notice until the 1960's. Literally hundreds of research efforts were made after that, but in foreign language testing the cloze test was still a relatively untried instrument up to the mid 1970's. In the past ten years, however, the cloze test has become increasingly popular with foreign language teachers. As Porter (1978) remarked, the cloze test is used as a "weapon in the teacher's armoury" (333). The cloze procedure has been used to measure readability, reading comprehension, language proficiency, aural and oral language ability, and reader's I.Q. My purpose in this paper is to review some aspects of using the cloze procedure in the measurement of reading comprehension and to reflect upon its use in the EFL classroom.

A good test possesses three characteristics: reliability, validity, and practicality (Harris 1969). The practicality of the cloze test is evident in the test's construction and scoring.

Cloze Test Construction

A cloze test usually consists of a prose passage in which some words are deleted for subsequent reconstruction by subjects. The most commonly employed deletion system in cloze studies is the "nth word deletion system," which systematically deletes words through a passage, for example, every eighth or ninth word. It seems evident that the cloze test is easy and quick to prepare and that no training or particular skills on the part of the test constructors are required.

FL 016 088

In preparing a cloze test for reading comprehension in an EFL class, the mechanical deletion system of nth word seems appropriate. Wainman's experiment (1979) showed that the mechanical deletion of every 10th word produces results that agree closely with results obtained by an intuitive deletion of content words. In his study, Wainman also noticed that EFL students' errors were made up almost equally of content and function words while first language learners' errors showed a much higher proportion of content errors. Since content and

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Christine Meloni

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

function words seemed to cause equal problems for ESL students, the cloze method of mechanical deletion could be expected to achieve results for ESL students which are as good as the results achieved by the deletion of the content words.

However, Anderson (1976) noted that the mechanical deletion system may produce some items that even skilled readers find ambiguous. It is very likely that these items are much too difficult or almost impossible for EFL students to reconstruct, while on the other hand, some very simple or non-discriminating items may be generated. Still, our purpose in using the random deletion system in constructing cloze tests for reading comprehension does not seem to be affected as materials contain different percentages of difficult and easy items, which are important factors in readability and are directly related to the testee's performance in reading comprehension. As it is explained by Taylor (1957) in defending the random-deletion system,

to restrict deletions to particular kinds of words is to ignore the fact that those kinds may not occur equally often in different materials. That difference in frequency of occurrence may itself be a readability factor. (22)

In using the random-deletion method, we must, however, make sure that the passages chosen are of a suitable length so as to allow for about 50 deletions. If a passage is too short, we may not be able to get the information we desire from our students.

In selecting materials for building the tests, I tend to think of using more "pre-cloze tests," that is, constructing cloze tests from materials students have not previously read in class, but possibly keeping a theme corresponding or related to the text being dealt with in class. Too many "post-cloze tests" (tests taken after the students have read or studied the un mutilated passage) may unintentionally encourage students to memorize the passages, thus reducing the reliability of using cloze tests as a measure for reading comprehension or for other purposes. Using pre-cloze tests inspires students to actively participate in the "psycholinguistic guessing game" - reading and comprehension.

Cloze Test Scoring

There are numerous ways of scoring the test, but the two most common ways are (1) to give credit for exact responses only and (2) to give credit for contextually-acceptable responses.

Although studies done by Stubbs and Tucker (1974) and many other researchers have shown significant and positive correlation ($r=.97$, $p<.01$) between scoring for exact versus contextually-acceptable responses, I still think it is worthwhile to score the test for contextually-acceptable answers in the EFL class. Allowing contextually-acceptable answers increases individual scores, which is encouraging to students. More importantly, contextually-appropriate responses are adequate to indicate students' comprehension of the materials being tested. To require verbatim answers, i.e. to restrict the correct response to exact word replacement only, seems to require more than merely comprehension.

In Hinofotis' study (1980), a cloze test, which was constructed from a passage of 427 words in which every 7th word was deleted up to a total of fifty deletions, was given to 107 subjects from a variety of native language backgrounds and with varying degrees of competence in English. The test was then scored both for the cloze-exact response and for the cloze-acceptable response. With the exact-word method, the average percent of correct responses was 24% and with the acceptable-word method, 31%. When the subjects were rank ordered according to both sets of scores, there were few differences in rankings, indicating the positive correlation between the two scoring methods. However, in the calculation of the standard deviations, noticeable differences were observed: 2.08 for the exact-word method and 7.3 for the acceptable-word method. This is not surprising since in the exact-word scoring method, no alternative answers were accepted. Thus, when a wide range of performance is expected as in a situation where students with widely varying language ability are tested, the acceptable-word scoring method is preferable as it provides more accurate information about the testees' levels of English. For EFL classroom use, the acceptable-word method is also more effective in that the information, i.e. the large amount of variance it provides, can serve as a very good diagnostic tool in assessing the students' understanding.

Results from certain item analyses (Harris lecture) have indicated that native speakers tend to use words which are more elaborate than the deleted words. For example, in item 2 of the following cloze test, only 21% of the 53 native English speakers being tested used the original word deleted - 'great.' The other 79% used other adjectives such as 'lasting,' 'significant,' 'profound,' 'drastic,' etc., which seem to fit perfectly in that context.

In the year 787 BEGAN a series of events WHICH
 were to have a GREAT effect upon the history OF the
 English language....

The above example suggests that the more competent users of the language have a large reservoir of vocabulary to choose from. Not using the original word deleted does not mean that they are not able to comprehend the context. This example also indicates the advantage of the contextually-acceptable scoring method over the exact-word scoring method in assessing the testees' language comprehension.

Reliability of the Cloze Test as a Measure of Reading Comprehension

The reliability of a test is manifested in the consistency of scores under repeated measurements.

In Anderson's study (1976), satisfactory reliability coefficients were obtained both by using a different scoring method for the same test and by using the test-retest method. Under different scoring measures (namely, the Verbatim Score Measure, in which each exact replacement of a deleted word was given one mark; the Synonym Score Measure, in which exact replacement of a deleted word was given one mark and each synonym was given half a mark; and the Alternative Response Score Measure, in which each exact replacement of a deleted word was given one mark and each grammatically correct response that fits the original syntactic pattern of language structure and makes sense within the context half a mark) reliability coefficients exceeding 0.90 were obtained for a 36-item test.

In the test-retest method, the same cloze reading comprehension test was administered to the same subjects. Depending on the scale of the test (the number of subjects taking part in the test and the number of items in the test), reliability coefficients from 0.78 to 0.96 were achieved (Anderson 1976). Besides English cloze tests, cloze test experiments in the Chinese and Malay languages were also conducted in a similar

way. Again, high reliability coefficients were obtained.

Porter (1978), however, questioned whether the cloze test is a reliable indicator since in his experiment two cloze tests, each with 50 blanks, constructed from identical material but beginning at two different and arbitrarily selected points, yielded markedly different sets of scores. The two tests were administered to a single population, with an interval of three weeks between the first and the second test. Disregarding the practice-effect, Porter found that the Pearson product moment coefficient of correlation was not as high as expected for a pair of tests which are so similar. For the exact word scores, the coefficient of correlation he obtained was 0.57 and for the acceptable-word scores 0.65. Porter argued that for cloze to be reliable as a test type, we should expect alternative and apparently equivalent forms of a test to show no significant difference in the results. He concluded that the relatively low correlations obtained with either scoring method indicate that students' achievement may vary markedly according to what is deleted and that the construction of cloze tests should be left to the professional tester.

Porter's results seem to be dramatically different from the results of the other studies cited above. As he took quite a different approach in his study, similar research may be needed to see if the results can be replicated, or an analysis of the two tests' cloze items could be conducted to find out exactly where the differences occur.

Validity of the Cloze Test for Reading Comprehension

The validity of a test can be determined if (1) the test is found to be based upon a sound analysis of the skill or skills to be measured, and (2) if there is sufficient evidence that test scores correlate fairly highly with actual ability in the skills being tested (Harris 1969).

To verify that cloze tests for reading comprehension meet the first criterion seems to be extremely difficult. First of all, there is not yet in existence a universally accepted definition of what language comprehension or reading skill is (Flahive 1980). Secondly, there is the issue raised by Farhady (1983) that the items in a cloze test are not developed by the test developer and therefore they cannot measure what the test developer wants to measure. Farhady asserts that if we do not know what a cloze test is measuring, we cannot establish its validity. However, a large number of criterion-related validity studies have

been carried out on cloze tests, from which, I hope, we can draw some insight.

The concept of internal and external information is relevant to a discussion of cloze tests as a measurement of reading comprehension. When Fanselow, in a classroom situation, showed a picture of camels, sheep, cattle, and shepherds, the students "saw" four legs on both camels even though the leg of one of the camels was actually obscured by a boy walking beside it. They also "saw" the heads of all the sheep and cattle although some of the heads were obscured by the shepherds. These illusions occurred because the viewers' internal information contributed to the external information provided in the picture itself. (In this case the external information is the visual message available and the internal information is the information we have in our heads based on our prior experience.)

It seems to me that reading comprehension requires, to a large extent, the integration of this internal and external information, and doing cloze tests is much like looking at Fanselow's partially obscured picture. In fact, this is similar to the concept of closure in Gestalt psychology, which refers to the natural human psychological tendency to fill in gaps in imperfect visual patterns. Porter (1976) remarked that the student, while being required to predict the missing words, is being tested on his ability to continuously draw from the context clues available and to draw information from his old knowledge to predict the nature of the language immediately ahead. This ability to predict what lies ahead depends on his ability to comprehend the language being processed and the internal information he possesses. This, to a certain degree, provides the justification for the cloze test as a test of comprehension.

Evidence of the validity of cloze tests as measure of readability was also shown in experiments in which the same test passages were administered to independent samples of subjects, and the samples ranked the passages in the same order of difficulty. In Anderson's experiment (1972), the cloze technique used with some Papua and New Guinea students ranked three passages of English prose in the same order of difficulty as did experienced reading teachers.

Anderson (1971) pointed out that, defined from a psycholinguistic viewpoint, reading comprehension is the correspondence between the semantic and grammatical habit system of encoders and decoders in the same language. If there is close correspondence between the decoder's and the encoder's system of language habits, then the passage or message is easy to comprehend. If there is not, comprehension is difficult. As

readability or reading difficulty of a passage and the reader's comprehension of a passage are not independent entities, I would think that the reliability of the cloze test as a measurement of readability lends itself to the validity of the cloze test as a measurement of reading comprehension. A student's performance on a cloze test gauges readability; readability in turn affects the student's performance on a cloze test.

Porter (1976) argued that language production is not comprehension. As the examinees are asked to generate the appropriate answer to fill the cloze item, they are being tested not only on their ability to comprehend but on their ability to produce the language as well. He concluded that it was misleading to suggest that the cloze test was measuring comprehension.

To measure what comprehension tests intend to measure, he suggested a modified cloze procedure which supplies a set of words for each cloze item. The modified cloze test has one great merit in that it has in it an aspect of control in supplying the sets of words for the cloze item. Whether it is a measure of pure comprehension is unknown, but it certainly takes a much longer time to prepare. Also, in the set of words for the cloze item, the choices among the distractors can show different degrees of relationship to the original answer, thus resulting in a scoring problem.

How should one evaluate the differences among the distractors for the cloze item? If we give credit to the correct answers only, we certainly miss a lot of information about how our students interpret the passage. It is obvious that when the modified cloze test combined the cloze procedure with the multiple-choice test, it inherited the merits and demerits of both.

For the EFL classroom the usual kind of cloze test as a measure of reading comprehension is recommended, as it is powerful and economical. Attempting to build tests which test purely reading comprehension but nothing else, may, in most cases, exceed the limits of EFL teachers' time and energy and may not be necessary. However, there is no harm in being aware that various aspects of linguistic and even non-linguistic ability may be involved in the performance of the cloze as a reading comprehension test. Stansfield and Hansen's experiment (1983) suggested there may be a cognitive bias operating in conjunction with cloze test performance; that is, a student's characteristic, field dependent-independent cognitive style, could be a variable in the cloze test performance.

Alderson (1983) also questioned the assumption that cloze tests are good devices for assessing reading comprehension. He found that the cloze items in the cloze tests he used tended to measure linguistic skills of a relatively lower order. According to Alderson, the cloze test is largely confined to the immediate environment of a blank because the procedure deletes single words rather than phrases or clauses.

As recognizing cohesive devices was defined as a higher-order skill related to reading comprehension, Brown (1980) investigated the problem of whether a cloze passage tests recognition of cohesive devices and to what degree. Cohesive devices such as those presented by Halliday and Hasan - reference, substitution, ellipsis, conjunction and lexical cohesion - were counted for the entire passage and then for the original blanks deleted. The results of these analyses indicated that such cohesive devices accounted for or were involved in between 56 and 70 percent of the items in different deletion patterns, while accounting for 52.25 percent of the words in the entire passage.

It appears from these results that cohesive devices play a rather substantial role in what the cloze passage is testing. Contrary to Alderson, Brown's findings confirmed that more than lower-order linguistic skills are involved in the cloze test and that the cloze test does test reading comprehension.

The validity of cloze tests as measures of reading comprehension is also determined by correlating cloze scores with those of standardized tests of reading and other measures of reading comprehension made over the same materials, which is also a way to try out the second criterion of Harris' test validity. In Anderson's experiment (1976), significant correlations between standard- and exact-length versions of cloze tests and a standardized measure of reading comprehension were obtained (0.61, 0.75 respectively). Other experiments on concurrent validity, predictive validity and construct validity all showed satisfactory results.

Compared with the traditional reading comprehension question tests and other types of comprehension tests, the cloze reading comprehension tests eliminate the comprehension problems posed in the question themselves, thus saving us the trouble of worrying about whether the student answered incorrectly because s/he could not understand the passage being tested or because s/he could not understand the questions.

Also, marking can be done by students soon after they have completed the tests, so that immediate knowledge of results and feedback to the teacher can be obtained. If the cloze test is followed by discussion, the students' comprehension of the original author might be enhanced. Another advantage of the cloze test is flexibility; the teacher can vary the word-blank frequency according to the ability level of the class or as the course progresses, although this may be effective only to a very limited extent.

In general, the cloze test is a practical and powerful measurement. Many results of criterion-related studies seem to be in favor of the cloze test as a test of reading comprehension, despite the fact that there still exists some controversy over its reliability and validity. Great efforts have been made in the research of cloze tests, but there is still much to be pursued, much to be confirmed or refuted. However, for EFL classroom use, the cloze reading comprehension test, as a component of testing, can play an important role and is expected to be used more extensively by EFL teachers.

THE AUTHOR

Wen-hui Li is a graduate student in the School of Languages and Linguistics at Georgetown University.

REFERENCES

- Alderson, J. Charles. 1983. The cloze procedure and proficiency in English as a foreign language. In Issues in language testing research, John W. Oller, Jr. (Ed.), 205-217.
- Anderson, Jonathan. 1971. A technique for measuring reading comprehension and readability. English Language Teaching 25: 178-182.
- Anderson, Jonathan. 1972. The application of cloze procedure to English learned as a foreign language in Papua and New Guinea. English Language Teaching 27: 66-72.
- Anderson, Jonathan. 1976. Psycholinguistic experiments in foreign language testing.

- Brown, James D. 1983. A close look at cloze: validity and reliability. In Issues in language testing research, John W. Oller, Jr. (Ed.), 237-250.
- Fanselow, John F. (forthcoming). Breaking rules: alternatives for language teachers.
- Farhady, Hossein. 1983. New directions for ESL proficiency testing. In Issues in language testing research, John W. Oller, Jr. (Ed.), 253-269.
- Flahive, Douglas E. 1980. Separating the g factor from reading comprehension. In Research in language testing, John Oller, Jr. and Kyle Perkins (Eds.), 34-46.
- Harris, David P. 1969. Testing English as a second language. New York: McGraw Hill.
- Hinofotis, Frances B. 1980. Cloze as an alternative method of ESL placement and proficiency testing. In Research in language testing, John W. Oller, Jr. and Kyle Perkins (Eds.), 121-126.
- Porter, D. 1976. Modified cloze procedure: a more valid reading comprehension test. English Language Teaching 30:152-155.
- Porter, D. 1978. Cloze procedure and equivalence. Language Learning 28:333-341.
- Stansfield, Charles and Hansen, Jacqueline. 1983. Field dependence-independence as a variable in second language cloze test performance. TESOL Quarterly 17 (1): 28-37.
- Stubbs, J.B. and Tucker, G.R. 1974. The cloze test as a measure of English proficiency. Modern Language Journal 50:239-241.
- Taylor, W.L. 1957. Cloze readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology 41:19-26.
- Wainman, H. 1979. Cloze testing of second language learners. English Language Teaching 33:126-132.